**5П**

**УДК 004.896:534.4**

***В.Е. Павловский, А.Ю. Кирков***
Институт прикладной математики имени М.В. Келдыша РАН
Россия, 125047, Москва, Миусская пл., 4

# Искусственный тональный язык акустической коммуникации роботов

***V.E. Pavlovsky, A.Ju. Kirkov***
*Keldysh Institute of Applied Mathematics of RAS (KIAM)*
*Russia, 125047, Moscow, Miusskaya sq., 4*

# *Artificial Tone Language for Acoustic Communication of Robots*

***В.Є. Павловський, А.Ю. Кирков***
Інститут прикладної математики ім. М.В. Келдиша РАН
Росія, 125047, Москва, Міуська пл., 4

# Штучна тональна мова акустичної комунікації роботів

Рассматривается задача коммуникации роботов на основе языка тональных акустических сигналов. Приведена формальная модель языка, в котором каждому символу языка соответствует последовательность акустических одночастотных сигналов, и рассмотрен алгоритм распознавания таких характерных звуков.
**Ключевые слова**: акустическая коммуникация, разговор роботов, распознавание речи

The task of communication of robots on the basis of tone acoustic signals is the focus of this work. In the work, the language model is presented where each symbol is associated with a pattern of single-frequency tones supplied with an algorithm of identification of these tones.
**Key Words**: acoustic communication, conversation of robots, speech recognition.

Розглядається задача комунікації роботів на ґрунті мови тональних акустичних сигналів. наведено формальну модель мови, в якій кожному символу мови відповідає послідовність акустичних одночастотних сигналів, і розглянуто алгоритм розпізнавання таких характерних звуків.
Ключові слова: акустична комунікація, розмова роботів, розпізнавання мовлення

## Введение

Цель работы – создание системы акустической коммуникации роботов для реализации их «разговора» в обычной внешней воздушной среде на акустических частотах, воспринимаемых человеком. Вообще говоря, малочастотный язык роботов не будет непосредственно понятен человеку, но для человека несложно подготовить соответствующий транслятор. При этом предлагаемая акустическая коммуникация может стать удобным средством дополнительного контроля поведения роботов, сосуществующих с человеком в едином социуме.

В настоящей работе коммуникация роботов основана на простых по своей структуре одно- и малочастотных сигналах, что делает методы спектрального анализа практически идеальными для последующих алгоритмов распознавания. На базе таких сигналов строится базовый язык роботов. В работе реализована программная среда для

визуализации и анализа работы построенного алгоритма. Проведены эксперименты на основе созданных компьютерной и аппаратной моделей для анализа работоспособности построенных алгоритмов. Первый прототип описываемой системы был создан в 2007 – 2008 гг., он кратко опубликован в сборнике и на электронных ресурсах [1]. В последующее время выполнялись эксперименты по созданию нескольких различных реализаций и отработке языка.

На рис. 1 показан мобильный робот «Аргонавт-2» семейства «Аргонавт», созданный в ИПМ им.М.В. Келдыша РАН. Робот оснащен приемной системой слуха, на этом роботе отрабатывались различные задачи акустической сигнальной обработки, в т.ч. базис (функции ввода, приема и передачи) описываемого языка.

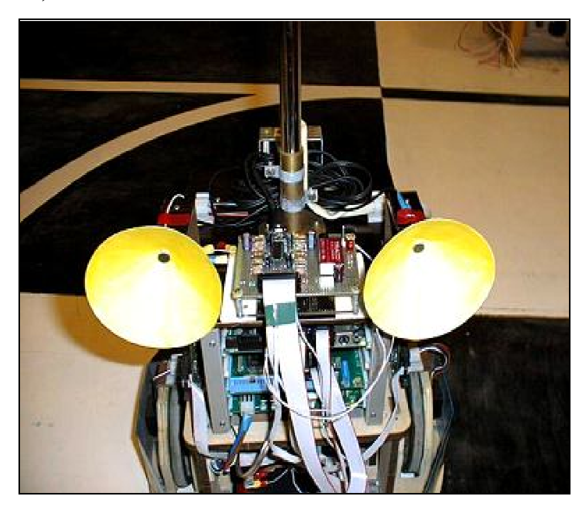



Рисунок 1 – Робот «Аргонавт-2» с системой слуха

Укажем, что в 2004 году группой американских специалистов была создана и испытана аналогичная система [2], в которой символы кодируются стандартными DTMF-сигналами (сигналы класса телефонных, когда телефон переключен в тональный режим). В 2011 году по этой же теме появилась публикация [3] авторов из австралийского университета Квинсленда (Queensland).

В отличие от указанных публикаций, в данной работе рассмотрена формальная модель языка. В этом языке символы могут кодироваться не только частотами DTMF, а любой последовательностью частот.

# 1 Структура языка роботов

Структура любого акустического языка состоит из алфавита и набора фонем, соответствующих каждой букве (символу) этого алфавита.

Фонема в данной работе делится на фрагменты. Фрагмент представляет собой произвольный одночастотный сигнал определенной длительности и амплитуды, т.е. элемент множества

$$Frg = \{(amp, frq, dur) \mid amp \in N, frq \in N, dur \in N\}$$

Фрагменты играют роль неделимых частиц, атомов языка. Из фрагментов составляются все другие конструкции языка. Поскольку фрагмент является одночастотным сигналом, будем называть систему одночастотной.

Длительность фрагмента $frg = (amp, frq, dur)$ будем обозначать $|frg| = dur$. Частоту фрагмента $frg = (amp, frq, dur)$ будем обозначать $frg[] = frq$.

Любое конечное подмножество

$$F \subset \{(frg^1, frg^2, ..., frg^n) \mid n \in N, frg^i \in Frg, i = \overline{1,n}\},$$

множества конечных упорядоченных наборов элементов из $Frg$ будем называть множеством фонем.

Таким образом, любая фонема представляет собой последовательность одночастотных сигналов. В отличие от фрагмента, фонема является многочастотным сигналом. Пример аудиосигнала, соответствующего фонеме, графически изображен на рис. 2, на котором фонема состоит из фрагментов трех частот.

Длительностью фонемы будем считать сумму длительностей всех входящих в неё фрагментов. Длительность фонемы $f = (frg^1, ..., frg^n)$ будем обозначать $|f| = \sum_{i=1}^{n} |frg^i|$. Для получения $k$-го фрагмента $frg^k$ фонемы $f = (frg^1, .., frg^k, .., frg^n)$ будем использовать запись $f[k] = frg^k$.

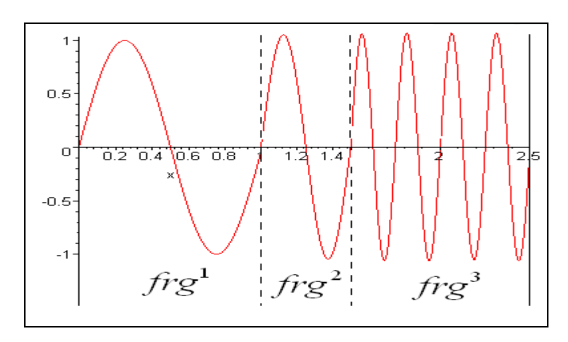



Рисунок2 – Пример аудиосигнала, соответствующего фонеме $f = (frg^1, frg^2, frg^3)$

Наложим на множество фонем $F$ несколько ограничений. Потребуем, чтобы ни одна фонема не являлась частью другой фонемы, т.е. для любой фонемы $f = (frg^1, ..., frg^n)$ не должно существовать такой фонемы $f' = (frg^{k_q}, ..., frg^{k_2})$, что $1 < k_1 < k_2 < n$. Данное условие будем называть ограничением вложенности фонем.

Помимо этого будем считать, что любая последовательность из двух фонем не содержит никакой фонемы. Т.е. ни одна фонема не является концом одной и началом другой фонемы. Иными словами, для любых двух фонем $f_1 = (frg_1^1, .., frg_1^{n_1}) \in F$ и $f_2 = (frg_2^1, .., frg_2^{n_1}) \in F$ не существует фонемы $f_3 \in F$, такой что

$$f_3 = (frg_1^{k_1}, .. frg_1^{n_1}, frg_2^1, .. frg_2^{k_2})$$

для некоторых $k_1, k_2 \in N$, $1 \le k_1 \le n_1$, $1 \le k_2 \le n_2$. Данное условие будем называть ограничением объединения фонем.

Введем также понятие множества используемых фрагментов, как набора всех фрагментов, входящих хотя бы в одну фонему.

$$Frg_u = \{frg \in Frg \mid \exists f = (frg^1, .., frg^k, .., frg^n), f \in F, frg^k = frg\}$$

Это множество конечно в силу конечности $F$.

Аналогично рассмотрим множество используемых частот, как набора частот, входящих хотя бы в один фрагмент из множества $Frg_u$

$$Frq_u = \{frq = frg[], frg \in Frg_u\}$$

Рассмотрим некоторое конечное множество $B$, которое будем называть алфавитом. Элемент этого множества будем называть буквой. Количество букв в алфавите, т.е. мощность множества $B$, обозначим $N_B$.

Множество фонем разобьем на $N_B$ непересекающихся непустых подмножеств $F_i$ и каждой букве поставим в соответствие одно из них. Это соответствие обозначим $R$. Таким образом, любой букве соответствует как минимум одна уникальная последовательность одночастотных сигналов.

Введем также специальный символ $S$, который будем называть пробелом. Поставим в соответствие этому символу некоторое число $dur_S \in N$, которое будем называть длительностью пробела. Пробел в рассматриваемом акустическом языке будет соответствовать тишине (сигналу с амплитудой ноль) длительностью не менее $dur_S$. Набор из символа пробела и его длительности обозначим $S^* = \{S, dur_S\}$.

Набор $L = (B, F, R, S^*)$ алфавита, множества фонем, соответствия между ними и пробела будем называть языком роботов.

Таким образом, робот, произнося слово из алфавита, преобразует его в последовательность фонем, что в свою очередь представляет собой последовательность одночастотных сигналов. Задача другого робота – выделить из всего звукового потока характерные звуки, соответствующие фонемам языка и на выходе получить слово в соответствии с распознанными фонемами.

## 2 Исходные понятия и схема распознавания

Целью алгоритма распознавания является поиск фонем в полученном роботом аудиосигнале. Как правило, робот на входе получает поток аудиоданных, которые сильно отличаются от исходных тональных сигналов. Пример аудиосигнала из эксперимента показан на рис. рис. 3. В связи с этим поставим задачу построить алгоритм, позволяющий из всего потока данных выделить звуки, соответствующие фонемам языка робота.

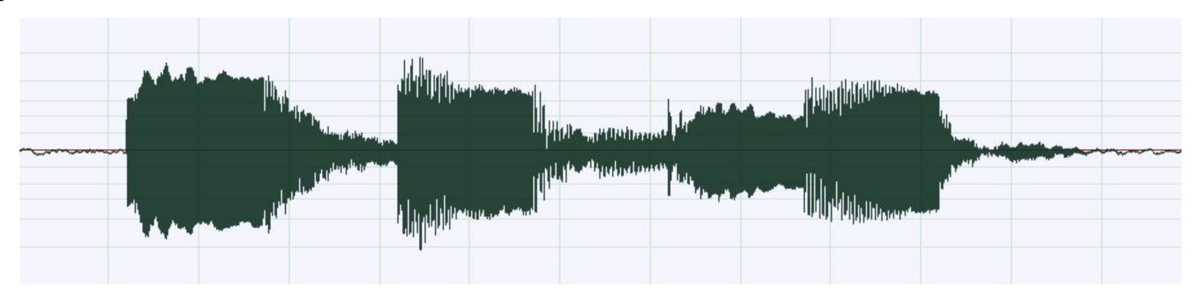


Рисунок3 – Пример аудио сигнала

Введем вначале ряд базовых понятий.

Аудиосигнал будем рассматривать как функцию $s(t)$, заданную на некотором отрезке $[t_1, t_2]$. Будем предполагать, что $s \in L_2([t_1, t_2])$. Задача распознавания заключается в определении всех интервалов $[t^{'}, t^{''}] \subset [t_1, t_2]$, на которых была проиграна некоторая фонема $f \in F$.

Базовый метод, который будем использовать, – метод спектрального анализа – позволяет на каждом интервале времени подсчитать амплитуду, соответствующую

частоте любого фрагмента. Используя формулы этого метода, рассмотрим следующую функцию

$$\underset{[t',t''],frq}{amp}\ s(t) = \sqrt{a^2 + b^2}, где\ a = \frac{1}{|[t',t'']|}\int_{t'}^{t''} s(t)\cos(2\pi \cdot frq \cdot t)dt \qquad (2.1)$$
$$b = \frac{1}{|[t',t'']|}\int_{t'}^{t''} s(t)\sin(2\pi \cdot frq \cdot t)dt,$$

которая на интервале $[t',t''] \subset [t_1,t_2]$ возвращает амплитуду аудиосигнала $s(t)$, соответствующую частоте $frq$.

Введем параметр $\tau$, значение которого будем рассматривать как параметр алгоритма. Будем выбирать $\tau$ таким образом, что длительности всех фрагментов $frg \in Frg_u$ будут кратными $\tau$. Такое значение обязательно найдется в силу конечности множества $Frg_u$.

Рассмотрим разбиение отрезка $[t_1,t_2]$ точками $t^0,t^1,...,t^{s-1},t^s$, где $t^i = t_1 + i\cdot\tau, i = \overline{0,s-1}$, $t^s = t_2, t^s - t^{s-1} \le \tau$, на отрезки $\Delta t^i = [t^{i-1},t^i], i = \overline{1,s}$. Множество точек разбиения отрезка $[t_1,t_2]$ обозначим $T$.

$$T = \{t^0,t^1,...,t^s\}, t^i = t_1 + i * \tau, i = \overline{0,s-1}, t^s = t_2, t^s - t^{s-1} \le \tau$$

На каждом отрезке $\Delta t^i, i = \overline{1,s}$ рассмотрим функцию $a_i(f)$, которая на этом отрезке будет определять громкость фонемы $f$. Данную функцию будем называть амплитудой фонемы на отрезке $\Delta t^i$.

Определим критерии вхождения фонемы в аудиосигнал на каждом отрезке $\Delta t^i, i = \overline{1,s}$. Для этого введем следующее определение.

Фонему $f \in F$ будем называть выделенной на отрезке $\Delta t^i, i = \overline{1,s}$ по параметрам $a_{\min} \in N, 0 < \sigma < 1$, если одновременно выполняются следующие условия:

1. $a_i(f) \ge a_{\min}$
2. $a_i(f) \ge \sigma \cdot \max\limits_{f\in F, f\notin E_i \cup N_i} a_i(f)$ (2.2)

Иными словами, фонему будем считать выделенной на отрезке времени $\Delta t^i, i = \overline{1,s}$, если на этом отрезке её амплитуда выше двух порогов: абсолютного и относительного. Абсолютный порог определяется первым условием (2.2) и задается некоторым значением $a_{\min}$. Относительный порог следует из второго условия (2.2) и равен произведению максимальной амплитуды среди всех фонем $f \in F$ на некоторое значение $\sigma \in (0,1)$. Значения $a_{\min}$ и $\sigma$ являются параметрами алгоритма.

Целью введения этих оценок (порогов) является фильтрация фонем. Абсолютный порог отсекает все «тихие» фонемы, тогда как относительный порог выделяет из «громких» фонем «наиболее громкие».

Вышеуказанные критерии будут являться основными условиями вхождения фонемы в аудиосигнал.

Множество выделенных фонем на отрезке $\Delta t^i, i = \overline{1,s}$ обозначим $S_i$. Отметим, что множества $S_i, i = \overline{1,s}$ зависят от функции амплитуды фонемы, которая пока не определена, т.е.

$$S_i = S_i'(a_i), i = \overline{1,s} \qquad (2.3)$$

Для определения функции амплитуды фонемы введем ряд дополнительных понятий.

Будем называть текущим временем распознавания фонемы $f \in F$ к моменту времени $t^i \in T, i = \overline{0,s}$ максимальный отрезок до $t^i$, на котором фонема $f \in F$ непрерывно входила в аудиосигнал, т.е. была выделенной на каждом из отрезков разбиения этого интервала. Текущее время распознавания фонемы характеризует, какое время фонема непрерывно проигрывалась до момента $t^i$. Нижеследующая формула определяет текущее время распознавания фонемы

$$T_i^C(f) = \begin{cases} T_{i-1}^C(f) + \tau, \ если\, f \in S_i \ и \ i \geq 1_i \\ 0, в \ остальных \ случаях \end{cases} \qquad (2.4)$$

Отметим, что текущее время распознавания фонемы к моменту времени $t^i$ зависит от текущего времени распознавания фонемы на предыдущей точке разбиения $T$ и от множества выделенных фонем $S_i$ на отрезке $\Delta t^i$.

$$T_i^C(f) = T_i^{C'}(f, T_{i-1}^C, S_i), i = \overline{1,s} \qquad (2.5)$$

Согласно формуле (1.4) к моменту времени $t^0$ текущее время распознавания $T_0^C(f)$ произвольной фонемы $f \in F$ равно нулю.

Используя понятие текущего времени распознавания фонемы, введем понятие текущего фрагмента фонемы на отрезке $\Delta t^i, i = \overline{1,s}$. Текущий фрагмент фонемы $f \in F$ будет характеризовать, какой именно фрагмент этой фонемы входит в аудиосигнал на отрезке $\Delta t^i, i = \overline{1,s}$ в том случае, если фонема $f$ входит в аудиосигнал. Иными словами, если фонема проигрывается, какой именно фрагмент проигрывается на отрезке $\Delta t^i$. Текущий фрагмент фонемы $f \in F$ на отрезке $\Delta t^i, i = \overline{1,s}$ будем обозначать $frg_C^i(f)$.

Для определения текущего фрагмента фонемы рассмотрим остаток от деления текущего времени распознавания фонемы на её длительность.

$$\tilde{T}_i^C(f) = \mathrm{mod}_{|f|}(T_i^c(f))$$

Определим функции $\tilde{T}_i^k(f), i = \overline{1,s}, k \in N$, которые для каждой фонемы $f \in F$, $f = (frg^1, .., frg^k, .., frg^n)$ на отрезке $\Delta t^i$ будут равны сумме длительностей всех фрагментов до $k$-го включительно.

$$\tilde{T}_i^k(f) = \sum_{j=1}^{k} |frg^j|$$

Таким образом, текущим фрагментом фонемы будем называть фрагмент под номером $k$, если $\tilde{T}_i^k(f) \geq \tilde{T}_i^C(f)$ и, либо $k = 1$, либо $\tilde{T}_i^{k-1}(f) < \tilde{T}_i^C(f)$. Ниже приводится общая формула для расчета текущего фрагмента.

$$frg_C^i(f) = f[k], \ где \ k = \min_{g(j) \geq 0}(j), \ g(j) = \tilde{T}_{i-1}^j(f) - \tilde{T}_{i-1}^C(f), \quad f = (frg^1, .., frg^k, .., frg^n), \ i = \overline{1,s} \qquad (2.6)$$

Отметим, что текущий фрагмент фонемы на отрезке $\Delta t^i$ полностью определяется из текущего времени распознавания этой фонемы к моменту времени $t^{i-1}$.

$$frg_C^i(f) = frg_C^{i\,'}(f, T_{i-1}^C),\ i = \overline{1,s} \quad (2.7)$$

Поскольку начало проигрывания фонемы может не совпадать с точками разбиения $T$, введем некоторую параметризованную функцию $\delta_i(f) \in [0,\tau), i = \overline{1,s}$, которая на каждом отрезке $\Delta t^i$ будет равна смещению начала проигрывания фонемы $f$ от ближайшей точки разбиения $T$ в том случае, если фонема $f$ проигрывается на этом отрезке. Введем обозначение отрезка разбиения со смещением.

$$\Delta t_\delta^i(f) = [t^{i-1} + \delta_i(f), t^i + \delta_i(f)],\ i = \overline{1,s}, f \in F$$

Теперь на основе введенных понятий определим функцию амплитуды фонемы. Будем считать амплитудой фонемы $f \in F$ на отрезке $\Delta t^i$ амплитуду частоты текущего фрагмента этой фонемы на отрезке $\Delta t_\delta^i(f)$, согласно формуле (2.1).

$$a_i(f) = \underset{\Delta t_\delta^i(f), frg_C^i[]}{amp} s(t), i = \overline{1,s} \quad (2.8)$$

Отметим, что текущая амплитуда фонемы была определена с использованием определения текущего фрагмента фонемы и функции смещения на отрезке $\Delta t^i$.

$$a_i(f) = a_i^{'}(f, frq_C^i, \delta_i),\ i = \overline{1,s} \quad (2.9)$$

Теперь определим критерий, который будет определять, была ли фонема проиграна на определенном отрезке времени. Для этого введем следующее определение.

Фонему $f \in F$ будем называть распознанной на отрезке $[t^k, t^l], t^k, t^l \in T$, $t^l - t^k = |f|$, если текущее время распознавания фонемы к моменту времени $t^l$ положительно и кратно длительности фонемы $f$, т.е. одновременно выполняются следующие два условия.

$$\begin{aligned} &1.\ T_l^C(f) > 0, \\ &2.\ \mathrm{mod}_{|f|}(T_l^c(f)) = 0 \end{aligned} \quad (2.10)$$

Данное определение будем рассматривать в качестве критерия вхождения фонемы в аудиосигнал.

Рассмотрим множества $R_i$, которые для каждого отрезка $\Delta t^i, i = \overline{1,s}$ будут содержать все фонемы $f \in F$, распознанные на некотором отрезке $[t^k, t^l], t^k, t^l \in T$, который содержит $\Delta t^i \subset [t^k, t^l]$. Множество $R_i$ будем называть множествами распознанных фонем на отрезке $\Delta t^i$.

Введем понятие распознанного пробела.

Пробел $S^* = \{S, dur_S\}$ будем называть распознанным на отрезке $[t^k, t^l], t^k, t^l \in T$, если одновременно выполняются следующие условия

$$\begin{aligned} &1.\ |t^l - t^k| \geq dur_S \\ &2.\ R_i = \varnothing,\ i = \overline{k+1, l} \\ &3.\ k > 0,\ l < s \\ &4.\ R_k \neq \varnothing \\ &5.\ R_{l+1} \neq \varnothing \end{aligned} \quad (2.11)$$

Это понятие будет определять те отрезки времени, на которых был проигран сигнал, соответствующий символу пробела.

Пробел будет распознан на тех отрезках времени, на которых не были распознаны фонемы и длительность которых не меньше длительности пробела. При этом справа и слева от такого интервала должны быть отрезки, на которых будет распознана некоторая фонема.

Полученные ключевые понятия распознанной фонемы и распознанного пробела зависят от функции $\delta_i(f)$, которая пока еще не определена. Рассмотрим теперь алгоритм определения начала фонемы и выведем формулу для $\delta_i(f)$.

# 3 Алгоритм определения начала фонемы

Поскольку начало проигрывания фонемы может не совпадать с точками разбиения $T$, поставим задачу определить смещение начала проигрывания фонемы относительно ближайшей точки разбиения слева. Для этого определим функцию $\delta_i(f)$, которая для каждого отрезка $\Delta t^i, i = \overline{1,s}$ будет совпадать с искомым смещением, если фонема была проиграна на этом отрезке.

В силу того, что фонема начинает проигрываться с первого фрагмента, задача поиска начала фонемы сводится к поиску начала проигрывания частоты первого фрагмента. Для этого рассмотрим параметризованную функцию $\delta_i^*(frq)$. Эта функция на отрезках $\Delta t^i, i = \overline{1,s}$ будет равна смещению начала проигрывания частоты $frq$ относительно ближайшей точки разбиения слева. Будем считать, что функция $\delta_i^*(frq)$ зависит исключительно от частоты $frq$, аудиосигнала $s(t)$ и параметров алгоритма.

Используя $\delta_i^*(frq)$, определим функцию смещения $\delta_i(f)$. Функция $\delta_i(f)$ для каждой фонемы $f \in F$ будет равна значению функции $\delta_i^*(frq)$ для частоты $frq$ первого фрагмента фонемы $f$ на тех отрезках, на которых фонема начинает проигрываться. Отрезки, на которых фонема $f$ начинает проигрываться, будут определяться следующим образом. К их началу текущее время распознавания фонемы $f$ будет равно нулю и амплитуда аудиосигнала, рассчитанная по формуле (2.1), на текущем отрезке разбиения $\Delta t^i$ или на следующем будет больше значения параметра $a_{\min}$. Если же фонема проигрывалась на предыдущем отрезке разбиения, т.е. текущее время распознавания этой фонемы к началу текущего отрезка разбиения больше нуля, вычислять значение $\delta_i^*(frq)$ нет необходимости. В этом случае значение функции $\delta_i(f)$ берется из предыдущего отрезка. В том случае, если не выполняется ни одно из вышеуказанных условий, функция $\delta_i(f)$ будет равна нулю.

Ниже приведена общая формула расчета функции $\delta_i(f)$.

$$\delta_i(f) = \begin{cases} \delta_i^*(frq),\ frq = frg[],\ frg = f[1],\ T_{i-1}^C(f) = 0\ u \\ \qquad \underset{frq,\Delta t^i}{amp}\ s(t) > a_{\min}\ или\ \underset{frq,\Delta t^{i+1}}{amp}\ s(t) > a_{\min} \\ \delta_{i-1}(f),\ T_{i-1}^C(f) > 0 \\ 0,\ T_{i-1}^C(f) = 0,\ \underset{frq,\Delta t^i}{amp}\ s(t) \le a_{\min},\ \underset{frq,\Delta t^{i+1}}{amp}\ s(t) \le a_{\min} \end{cases} \tag{3.1}$$

Отметим, что функция $\delta_i(f)$ зависит от текущего времени распознавания фонемы к моменту времени $t^{i-1}$.

$$\delta_i(f) = \delta_i^{'}(f, T_{i-1}),\ i = \overline{1,s} \qquad (3.2)$$

Однако для полного определения функции $\delta_i(f)$ необходимо ввести формулу расчета функции $\delta_i^*(frq)$. Ниже приводится алгоритм расчета функции $\delta_i^*(frq)$.

Данный алгоритм будет заключаться в поиске первого отрезка длительностью, равной длительности отрезка разбиения $\tau$, на котором полностью была проиграна искомая частота.

На рис. 4 показан график изменения амплитуды, соответствующей частоте первого фрагмента фонемы из эксперимента. На графике амплитуда вычисляется через каждые 5 мс согласно формуле (2.1) по интервалу длительности 50 мс.

На графике начало проигрывания фонемы указано вертикальной стрелкой. Сплошной жирной чертой отмечена ближайшая точка разбиения. Область графика справа от стрелки (между двумя пунктирными линиями) соответствует периоду, в течение которого фонема была проиграна. Сильное увеличение амплитуды до начала проигрывания фонемы, объясняется тем, что интервалы, на которых считается эта амплитуда (50 мс), включает в себя часть периода проигрывания фонемы. Область графика после периода проигрывания фонемы соответствует эхо, которое является следствием отражения звуковых волн. Значение $\delta^*$ между точкой разбиения (сплошная линия) и началом проигрывания фонемы (точка, отмеченная стрелкой) является искомым смещением.

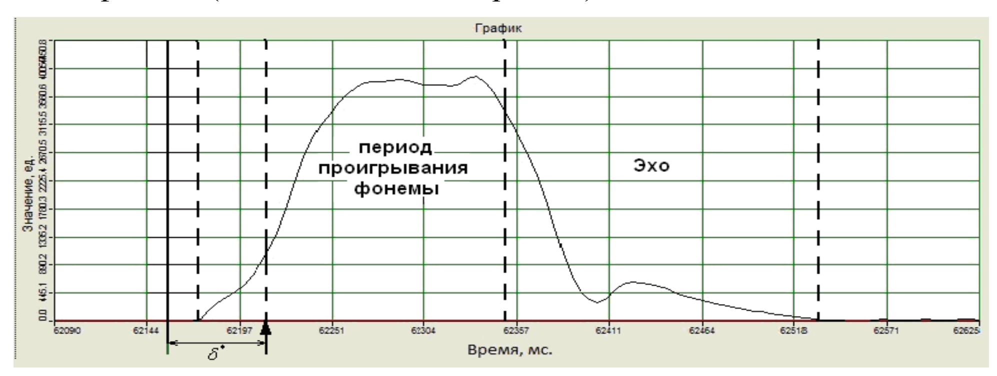



Рисунок4 – График изменения амплитуды

Для вычисления значения $\delta^*$ для произвольной частоты $frq$ на отрезке $\Delta t^i$ рассмотрим следующие интервалы

$$\Delta t_\delta^i = [t^{i-1} + \delta, t^i + \delta],\ \delta \in (0,\tau),\ i = \overline{1,s}$$

Определим функцию $amp_i(frq,\delta)$, которая для каждого $\delta \in (0,\tau)$ будет равна амплитуде частоты $frq$ на отрезке $\Delta t_\delta^i$ согласно формуле (2.1)

$$amp_i(frq,\delta) = \underset{\Delta t_\delta^i, frq}{amp}\ s(t),\ \delta \in (0,\tau),\ i = \overline{1,s}$$

Рассмотрим некоторый параметр $\alpha \in (0,1)$, значение которого будет являться параметром алгоритма. Определим наименьшее значение $\delta^*$, для которого будет выполнено неравенство

$$amp_i(frq,\delta^*) \geq \alpha * \max_{\delta\in(0,1)} amp_i(frq,\delta)$$

Значение $\delta^*$ будет являться искомым смещением. Ниже приведена общая формула для расчета функции $\delta_i^*(frq)$

$$\delta_i^*(frq) = \min_{g(\delta)\geq 0,\delta\in(0,1)}(\delta) \;, где \; g(\delta) = amp_i(frq,\delta) - \alpha * \max_{\delta'\in(0,1)} amp_i(frq,\delta') \qquad (3.3)$$

Отметим, что согласно вышеуказанной формуле, $\delta_i^*(frq)$ зависит только от частоты $frq$, аудиосигнала $s(t)$ и параметров алгоритма. Таким образом, ограничения, наложенные на функцию $\delta_i^*(frq)$, выполняются.

# 4 Проверка корректности формул

Ряд введенных функций и множеств в данной работе были определены «в рекуррентном стиле» с использованием понятий, которые в свою очередь зависят от этих же функций и множеств. Например, для введения множества выделенных фонем $S_i$ использовались функции амплитуды фонемы $a_i(f)$, текущего фрагмента фонемы $frg_C^i(f)$ и текущего времени распознавания $T_i^C(f)$, которые в свою очередь были определены с использованием множества выделенных фонем.

В связи с вышесказанным, необходимо проверить корректность введенных определений. Для проверки корректности покажем, что зная исходные данные на отрезке $\Delta t^i$ и всех предыдущих, можно вычислить все функции и множества для этого отрезка, а также получить исходные данные на отрезке $\Delta t^{i+1}$.

В качестве исходных данных на отрезке $\Delta t^i$ будем рассматривать функцию текущего времени распознавания фонемы $T_{i-1}^C(f)$. В силу того, что при $i=1$ эта функция $T_0^C(f)=0$ известна, указанная проверка будет доказывать корректность введенных определений. Последовательно покажем, что зная исходные данные на отрезке $\Delta t^i$ и предыдущих, можно определить функцию текущего фрагмента фонемы $frg_C^i(f)$, функцию смещения фонемы $\delta_i(f)$, функцию амплитуды фонемы $a_i(f)$, множества выде-ленных фонем $S_i$ и исходные данные на отрезке $\Delta t^{i+1}$, т.е. функцию текущего времени распознавания фонемы $T_i^C(f)$.

Согласно (2.7), (3.2) текущий фрагмент фонемы $frg_C^i(f)$ и функция смещения $\delta_i(f)$ уже зависят от исходных данных.

$$frg_C^i(f) = frg_C^{i\,'}(f, T_{i-1}^C),\; i=\overline{1,s}$$
$$\delta_i(f) = \delta_i^{'}(f, T_{i-1}),\; i=\overline{1,s}$$

Последовательно используя равенства (2.9), (2.7), (3.2) получаем, что функция амплитуды фонемы на отрезке $\Delta t^i$ зависит только от текущего времени распознавания фонемы к моменту $t^{i-1}$.

$$a_i(f) = a_i^{'}(f, frq_C^i, \delta_i) = a_i^{''}(f, T_{i-1}^C),\ i = \overline{1,s} \tag{4.1}$$

Согласно (2.3) и (4.1) получаем равенство, которое доказывает справедливость определения множества выделенных фонем

$$S_i = S_i^{'}(a_i) = S_i^{''}(T_{i-1}), i = \overline{1,s} \tag{4.2}$$

Таким образом, на отрезке $\Delta t^i$ доказана зависимость всех функций и множеств исключительно от функции текущего времени распознавания фонемы к началу этого отрезка. Покажем теперь, что зная эту функцию к началу отрезка $\Delta t^i$, можно получить значения аналогичной функции к концу этого отрезка.

Согласно (2.5), (4.2) получаем равенство, которое доказывает это утверждение.

$$T_i^C(f) = T_i^{C'}(f, T_{i-1}^C, S_i) = T_i^{C''}(f, T_{i-1}^C), i = \overline{1,s} \tag{4.3}$$

Таким образом, показана зависимость исходных данных на отрезке $\Delta t^{i+1}$ от исходных данных на отрезке $\Delta t^i$ и предыдущих. Следовательно, доказана корректность определения всех функций и множеств, рассмотренных в данной работе.

# 5 Алгоритм распознавания

Процесс обработки аудиосигнала (распознавания) начинается с прослушивания аудиоданных длительностью $2\tau$. Такой отрезок необходим для работы алгоритма определения начала фонемы. Далее аудиосигнал будем обрабатывать через каждый интервал длительностью $\tau$. На рис. 5 текущий интервал отмечен более жирной чертой, менее жирной чертой отмечены интервалы, которые прослушаны к этому моменту времени.

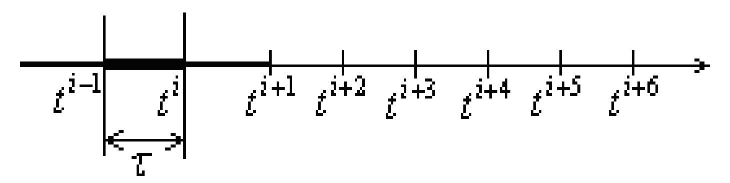



Рисунок5 – Временные интервалы распознавания

В качестве исходных данных на каждом отрезке $\Delta t^i$ будем рассматривать значения функции $T_{i-1}^C(f)$ текущего времени распознавания фонемы к началу этого отрезка. Отметим, что на первом отрезке исходные данные, $T_0^C(f) = 0$, $f \in F$, известны.

Обработка аудиосигнала на каждом отрезке $\Delta t^i$ будет заключаться в определении текущего времени распознавания $T_i^C(f)$ каждой фонемы $f \in F$ к концу этого отрезка.

Рассмотрим процесс обработки аудиосигнала на произвольном отрезке $\Delta t^k$. Как отмечалось, к началу этого отрезка нам известна функция $T_{k-1}^C(f)$.

Согласно определению для вычисления текущего времени распознавания $T_k^C(f)$ фонемы $f \in F$ к концу отрезка необходимо проделать ряд следующих действий.

1. Определить текущий фрагмент $frg_C^k(f)$ каждой фонемы $f \in F$.

2. Вычислить значение функции смещения $\delta_k(f)$ для каждой $f \in F$.

3. Вычислить значение функции амплитуды фонемы $a_k(f)$ для каждой $f \in F$.

4. Определить выделенные фонемы $S_k$.

Ниже представлено детальное описание каждого из вышеуказанных шагов.

*Определение текущего фрагмента.*

Текущий фрагмент $frg_C^k(f)$ фонемы $f \in F$ на отрезке $\Delta t^k$ будем определять согласно формуле (2.6). Для этого вычислим значение следующей функции.

$$\widetilde{T}_{k-1}^{C}(f) = \mathrm{mod}_{|f|}(T_{k-1}^{c}(f))$$

Далее определим наименьшее значение $l \in N$, для которого значение функции

$$\widetilde{T}_{k-1}^{l}(f) = \sum_{j=1}^{l} | f[j] |$$

будет не меньше значения функции $\widetilde{T}_{k-1}^{C}(f)$. Фрагмент $f[l]$ фонемы $f$ под номером $l$ будем считать текущим фрагментом на отрезке $\Delta t^k$.

$$frg_C^k(f) = f[l]$$

*Определение функции смещения.*

Определив текущий фрагмент фонемы $f$, переходим к вычислению значения функции смещения $\delta_k(f)$.

Согласно формуле (3.1), если $T_{k-1}^{C}(f) > 0$, в качестве значения функции $\delta_k(f)$ берется значение $\delta_{k-1}(f)$ на предыдущем шаге.

В том случае, если амплитуда частоты текущего фрагмента фонемы $f$, вычисленная согласно формуле (2.1), на отрезках $\Delta t^k$ и $\Delta t^{k+1}$ не превосходит значение параметра $a_{\min}$, считаем $\delta_k(f) = 0$.

В противном случае $\delta_k(f)$ будет равно значению функции $\delta_k^*(frq)$ для частоты $frq$ текущего фрагмента фонемы $f$.

Для вычисления $\delta_k^*(frq)$ методом стандартных отрезков рассмотрим интервалы

$$\Delta t_j^k = [t^{k-1} + j * \tau / 10, t^k + j * \tau / 10], j = \overline{0,9}$$

и вычислим амплитуду аудиосигнала, соответствующую частоте $frq$ текущего фрагмента $frg_C^k(f)$ фонемы $f$, на каждом из этих интервалов.

$$amp_j = \underset{\Delta t_j^k, frg^1}{amp} s(t), j = \overline{0,9}$$

Далее вычислим максимальное значение $amp_j$ амплитуды на этих интервалах и определим наименьшее значение $j^*$, для которого будет выполнено неравенство

$$amp_{j^*} \ge \alpha * \max_{j=\overline{0,9}} \underset{\Delta t_j^k, frg^1}{amp} s(t)$$

Значение $\delta = j^* * \tau / 10$ будет являться искомым смещением.

*Определение амплитуды фонемы*

Определив текущий фрагмент $frg_C^k(f)$ и значение функции смещения $\delta_k(f)$ для фонемы $f$, переходим к вычислению значения функции амплитуды фонемы $a_k(f)$. Согласно формуле (2.8) $a_k(f)$ будет равна амплитуде аудиосигнала, соответствующей частоте текущего фрагмента этой фонемы, вычисленной на отрезке со смещением $\Delta t_\delta^k(f) = [t^{k-1} + \delta_k(f), t^k + \delta_k(f)]$, используя формулу (2.1).

$$a_k(f) = \underset{frg_C^k(f)[], \Delta t_\delta^k(f)}{amp} s(t)$$

*Определение множества выделенных фонем.*

Вычислив значение функции амплитуды $a_k(f)$ каждой фонемы $f \in F$, переходим к определению множеств выделенных фонем $S_k$. Для этого, используя множества $E_k$, $N_k$, проверяем условия (2.2) определения выделенной фонемы для каждой $f \in F$. В том случае, если указанные условия удовлетворяются, считаем $f \in S_k$.

Множество $S_k$ после работы алгоритмов корректировки смещения влево и корректировки смещения направо будем считать итоговым множеством выделенных фонем. Используя это множество, определим текущее время распознавания $T_k^C(f)$ каждой фонемы $f \in F$ к концу отрезка $\Delta t^k$.

Получив множество выделенных фонем, переходим к определению текущего времени распознавания фонемы $T_k^C(f)$. Согласно формуле (2.4) текущее время распознава-ния $T_k^C(f)$ фонемы $f \in F$ будет равно значению $T_{k-1}^C(f)$ на предыдущем отрезке плюс длина отрезка разбиения $\tau$ в том случае, если $f$ является выделенной фонемой, т.е. $f \in S_k$, и нулю, в противном случае.

Определив текущее время распознавания $T_k^C(f)$ каждой фонемы $f \in F$ к концу отрезка $\Delta t^k$, переходим поиску распознанных фонем.

Проверяем условия (2.10) определения распознанной фонемы. В том случае, если для некоторой фонемы $f \in F$ условия удовлетворяются, считаем фонему $f$ распознанной на интервале $[t^k - |f|, t^k]$.

Будем обозначать $\Delta t^{R(i)} = [t_1^{R(i)}, t_2^{R(i)}]$ интервалы, на которых были распознаны фонемы. Значение $i$ будем определять согласно очередности. Т.е. при определении пер-вого отрезка, на котором распознана некоторая фонема, значение $i$ будет равно единице, второго – двум и так далее. Отрезки $\Delta t^{R(i)}$ будем называть распознанными.

После определения второго и последующих интервалов, на которых некоторая фонема будет распознанной, будем рассматривать отрезок $\Delta t^{S(i)} = [t_2^{R(i)}, t_1^{R(i+1)}]$ от конца предыдущего распознанного отрезка до начала текущего. Данный отрезок проверяется на удовлетворение условий (2.11) определения распознанного пробела. В том случае, если условия удовлетворяются, пробел считается распознанным на отрезке $\Delta t^{S(i)}$.

# 6 Калибровка

Как правило, частота, воспроизводимая одним роботом, отличается от частоты, которую воспринимает другой робот. На рис. 6 представлен график соответствия этих частот в проведенных экспериментах.

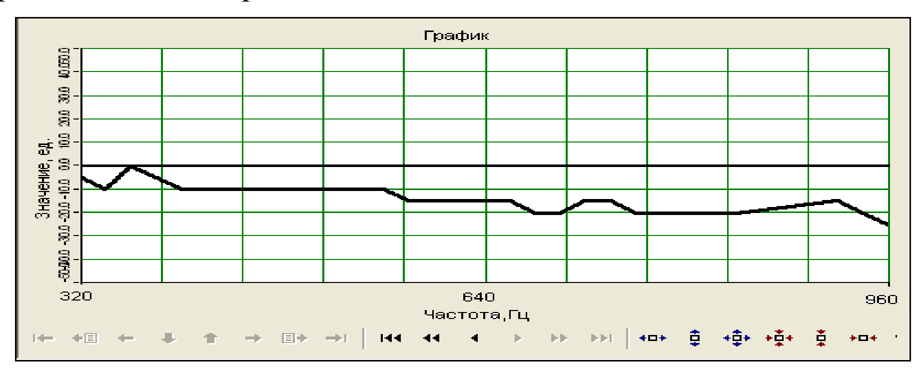



Рисунок6 – Изменение частоты в «передающем тракте»

В алгоритме для более качественного распознавания ставится задача получения соответствия между этими частотами. Определим функцию связи частот

$$fqr^* = K(frq)$$

Данная функция будет, принимая значение частоты, которую проигрывает один робот, выдавать значение частоты, которую слышит другой. Эта функция определяется опытным путем и используется в алгоритме распознавания.

# 7 Анализ работы алгоритма

Для реализации алгоритма разработана программная среда. При создании программной среды использовались методы, описанные в [4]. Программная среда позволяет осуществлять и анализировать работу рассматриваемого алгоритма для системы, состоящей из двух роботов.

Для тестирования системы в качестве алфавита рассмотрено множество букв русского алфавита. Каждой букве поставлено в соответствие множество, состоящее из одной фонемы, которая в свою очередь состоит из двух фрагментов, одинаковой длительности и амплитуды.

Таким образом, каждой букве алфавита соответствует двухчастотный сигнал. Частоты фрагментов, которые используются для каждой буквы, а также амплитуда и длительность этих фрагментов, приведены в таблице табл. 1. Все параметры отобраны в результате нескольких серий тестов и в этом смысле являются наилучшими.

Эти параметры вместе с выбором языка роботов подобраны экспериментальным путем и отвечали наилучшему качеству распознавания среди всех рассмотренных здесь параметров.

В качестве входного аудиосигнала в программе используется 8-битный моносигнал с частотой дискретизации 12800 Гц. Выбор этих параметров соответствует техническим характеристикам используемых в системе динамиков, микрофонов и аналого-цифровых преобразователей (АЦП).

Анализ работы алгоритма заключался в проведении нескольких тестов. Тестирование проводилось на двух компьютерах, один из которых («проигрывающий») использовалась для проигрывания звуковых фрагментов, а другой («распознающий») – для выделения из пришедшего потока аудиоданных звуков, соответствующих буквам алфавита.

Таблица 1 – Соответствие между буквами алфавита и параметрами сигнала

| Буква | Амплитуда | Частота 1-го фрагмента, Гц | Частота 2-го фрагмента,Гц | Длительность, Мсек |
|---|---|---|---|---|
| а | 100 | 400 | 2000 | 150 |
| б | 100 | 450 | 2050 | 150 |
| в | 100 | 500 | 2100 | 150 |
| г | 100 | 550 | 2150 | 150 |
| д | 100 | 600 | 2200 | 150 |
| е | 100 | 650 | 2250 | 150 |
| ж | 100 | 700 | 2300 | 150 |
| з | 100 | 750 | 2350 | 150 |
| и | 100 | 800 | 2400 | 150 |
| й | 100 | 850 | 2450 | 150 |
| к | 100 | 900 | 2500 | 150 |
| л | 100 | 950 | 2550 | 150 |
| м | 100 | 1000 | 2600 | 150 |
| н | 100 | 1050 | 2650 | 150 |
| о | 100 | 1100 | 2700 | 150 |
| п | 100 | 1150 | 2750 | 150 |
| р | 100 | 1200 | 2800 | 150 |

Продолжение табл. 1

| | | | | |
|---|---|---|---|---|
| с | 100 | 1250 | 2850 | 150 |
| т | 100 | 1300 | 2900 | 150 |
| у | 100 | 1350 | 2950 | 150 |
| ф | 100 | 1400 | 3000 | 150 |
| х | 100 | 1450 | 3050 | 150 |
| ц | 100 | 1500 | 3100 | 150 |
| ч | 100 | 1550 | 3150 | 150 |
| ш | 100 | 1600 | 3200 | 150 |
| щ | 100 | 1650 | 3250 | 150 |
| ъ | 100 | 1700 | 3300 | 150 |
| ы | 100 | 1750 | 3350 | 150 |
| ь | 100 | 1800 | 3400 | 150 |
| э | 100 | 1850 | 3450 | 150 |
| ю | 100 | 1900 | 3500 | 150 |
| я | 100 | 1950 | 3550 | 150 |

Параметры алгоритма:

– $\tau = 50$

– $a_{\min} = 32$

– $\sigma = 0.9$

– $\alpha = 0.7$

*Тест №1.* На разных расстояниях от микрофона «распознающей» машины динамиком «проигрывающей» машины один раз проигрывался звуковой сигнал, соответствующий фразе «абвгдежзийклмнопрстуфхцчшщъыьэюя». После чего результат распознавания сравнивался с исходной фразой. Результаты эксперимента представлены в таблице табл. 2

Таблица 2 – Результаты эксперимента №1

| Расстояние, м. | Распознанная фраза |
|---|---|
| 0.5 | абвгдежзийклмнопрстуфхцчшщъыьэюя |
| 1 | абвгдежзийклмнопрстуфхцчшщъыьэюя |
| 1.5 | абвгдежзийклмнопрстуфхцчшщъыьэюя |
| 2 | абвгдежзийклмнопрстуфхцчшщъыьэюя |
| 2.5 | абвгдезийклмнопстуфхцчшщъыьэюя |
| 3 | абвгдезийкмнопрстфхцчшщъыьэюя |
| 3.5 | абвгдезийклмнопрстуфхцчшъыьэюя |
| 4 | абвгдежзийклнопрстуфхцчшъыьэюя |
| 4.5 | абвгдезийклмнопрстуфхцчшъыьэюя |
| 5 | абвгдезийлмнопрстуфхчшъыьэюя |
| 6 | абвгдезиймнопрстуфхчшъыьэюя |
| 7 | абвгдезийкнопрстуфхчшъыьэюя |
| 8 | абвгдезийклнпрстуфхчшъьэюя |
| 9 | абвгдезийклнпсуфхшщъьэя |
| 10 | абвгдезийклнпсуфхшщъэя |

На рис. 7 изображен график зависимости процента распознавания от расстояния между коммуницирующими машинами.

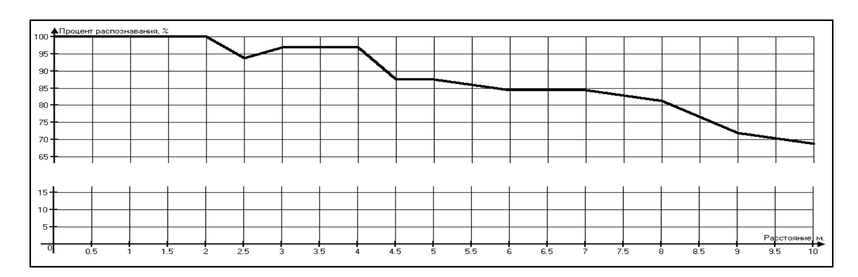



Рисунок7 – Процент распознавания как функция расстояния. Тест 1

*Тест №2.* На расстоянии двух метров от микрофона «распознающей» машины динамиком «проигрывающей» машины десять раз проигрывался звуковой сигнал, соответствующий фразе «абвгдежзийклмнопрстуфхцчшщъыьэюя». На рис. 8 изображен график зависимости процента удачного распознавания от номера эксперимента.

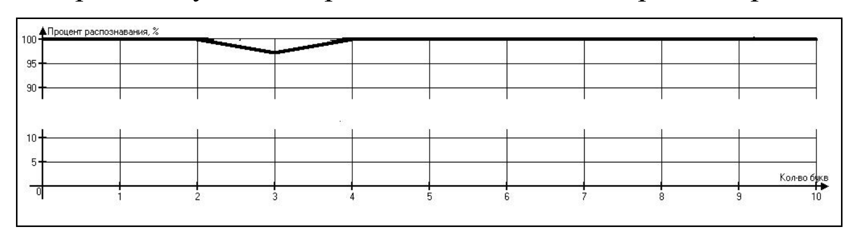



Рисунок8 – Тест повторного проигрывания-распознавания. Тест 2

*Тест №3.* На расстоянии двух метров от микрофона «распознающей» машины динамиком «проигрывающей» машины проигрывался звуковой сигнал, соответствующий разному количеству букв «а». График зависимости процента удачного распознавания от количества букв «а» представлен на рис. 9.

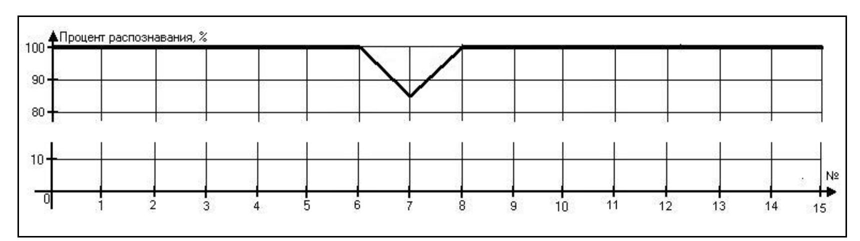



Рисунок9 – Тест повторного проигрывания-распознавания. Тест 3

Укажем здесь, что данные тесты выполнялись неоднократно, в разных состояниях внешней среды (температура, соответствующее давление воздуха, небольшая зашумленность). Во всех экспериментах результаты получались идентичными приведенным. Как видно из тестов на расстоянии до двух метров распознавание четкое (надежное), а далее возникают небольшие отклонения.

# 8 Схема использования языка с подкрепляющим диалогом

Рассмотрим следующий модельный пример применения системы тональной акустической коммуникации. На площадке квадратной формы, разбитой на 64 ячейки как шахматная доска, находятся два робота. Каждая ячейка окрашена некоторым цветом (рис. 10).

Каждый робот может передвигаться по ячейкам площадки и способен определить цвет ячейки, на которой он находится в данный момент. Роботы общаются между собой, используя тональный акустический язык.

Передвигаясь по ячейкам площадки, роботы сообщают друг другу информацию о цвете ячеек. При этом они используют следующую терминологию. Оказавшись на новой, неисследованной ячейке робот говорит фразу «Нахожусь в ячейке с координатами *X Y* цвет *такой-то*». При коммуникации существенно используется словарь возможных слов, описывающих координаты (целые числа) и цвета. В случае если второй робот «понял» первого, он ничего не отвечает. В обратном случае второй робот произносит фразу «не понял повтори» и первый должен повторить сказанную фразу.

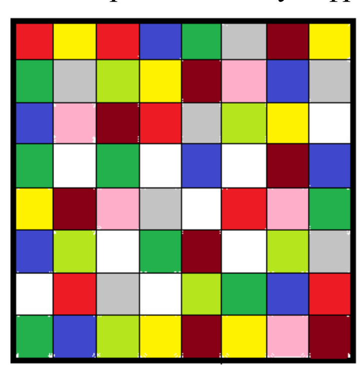


Рисунок10 – Площадка для исследования роботами

Рассмотрим результаты эксперимента в тех случаях, когда роботы находятся в ячейках на расстояниях 2 и 4 метра. Робот, находясь в ячейке с координатами (2, 1), говорит следующую фразу: «нахожусь в ячейке два один цвет желтый». График и спектрограмма исходного сигнала, соответствующего данной фразе, изображены на рис. 11. На графике изображены чистые гармоники с очень большой периодичностью.

Второй робот, находясь на расстоянии два метра, слушает этот сигнал. На рис. 12 изображен аудиосигнал, который принял второй робот. В результате эксперимента второй робот распознал фразу полностью.

Далее роботы находятся на расстоянии четыре метра, и первый робот произносит ту же самую фразу. Второй робот слушает первого и получает аудиосигнал, который изображен на рис. 13. Фраза, которую распознал второй робот следующая «нахожусь в ячейке два дин цвет желты». Фраза распознана не полностью, правильная фраза по-прежнему «нахожусь в ячейке два один цвет желтый», и второй робот просит первого повторить, произнося фразу «не понял повтори». Первый робот вновь произносит фразу, и теперь второй робот распознает её полностью.

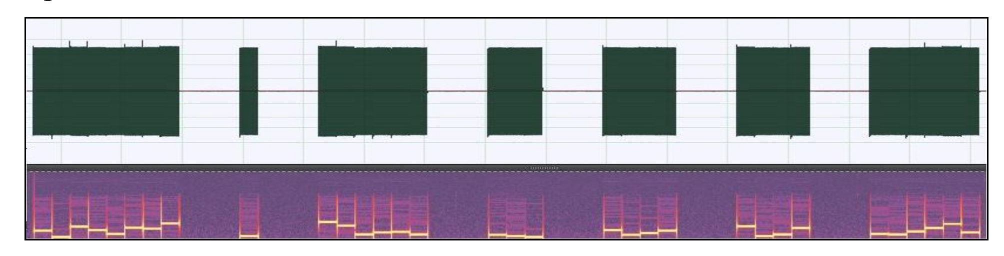


Рисунок11 – Исходный аудиосигнал

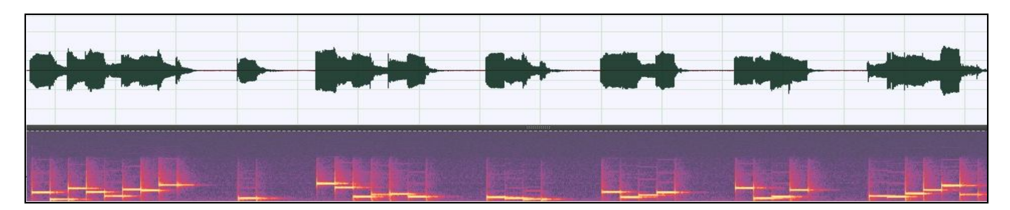


Рисунок12 – Аудиосигнал, полученный с расстояния 2 метра

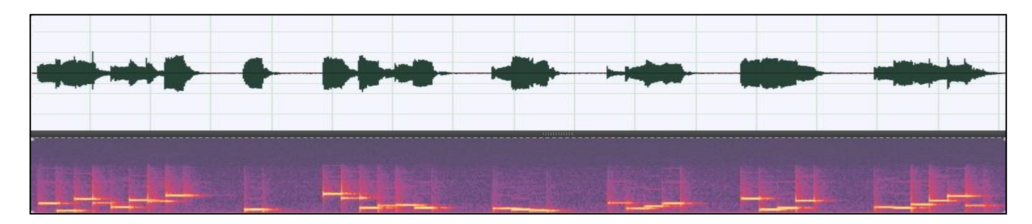


Рисунок13 – Аудиосигнал, полученный с расстояния 4 метра

Рассмотренный практический пример реализации диалога двух роботов на базе тонального акустического языка и диалог с подтверждением повышают надежность передачи и распознавания информации.

# Заключение

В работе выполнено построение модели акустического тонального языка роботов и разработан алгоритм выделения звуков, соответствующих фонемам этого языка. Данная система позволяет осуществлять коммуникацию роботов на основе акустических сигналов. В основе метода распознавания лежит программная реализация спектрального анализа аудиосигнала.

Выполнена реализация программной среды моделирования. Программная среда позволяет определять и анализировать параметры, характеризующие аудиосигнал. Эти параметры в дальнейшем будут использоваться для доработки данного алгоритма с целью достижения лучшего качества распознавания.

Моделирование показало работоспособность созданной схемы управления. Ее отработка будет продолжена на макетах подвижных роботов.

Отметим, что передача акустических сигналов во многих отношениях, таких как надежность, скорость, уступает передаче данных по радиоканалу. Однако данная система может использоваться в тех случаях, где неприменимы радиосигналы, например, в условиях значительных радиопомех. Также она может быть интересна при использовании в таком окружении, когда коммуникация роботов должна быть наглядной.

В дальнейшем планируется создать версии акустического языка, в которых каждый символ будет кодироваться не одно-, а двух- и/или трехчастотным сигналом. Цель – создание таких языков, которые будут более устойчивы к помехам.

Также планируется ввести некоторый понятийный словарь (тезаурус), которым будут пользоваться роботы. Для этого словаря будут созданы правила составления предложений, каждое из которых будет уже иметь какой-то смысл. Алгоритмы распознавания будут дополнены анализом этих предложений. Будет создан инструмент для пополнения данного словаря. Таким образом, на основе такого языка можно будет осуществить смысловой диалог роботов.

# Литература

1. Динамика, моделирование, управление мобильными роботами / [Павловский В.Е., Евграфов В.В., Павловский В.В., Петровская Н.В.] // Искусственный интеллект – проблемы и перспективы. Политехнические чтения (3. Поспеловские чт-я) : сборник статей. – М.м: Изд. Политехнического музея, – 2008.; Режим доступа : http://posp.raai.org/info.php?view=17
2. Tho Nguyen and Linda G.Bushnell, Dept of EE, University of Washington, April, 2006. “Feasibility Study of DTMF Communications for Robots” [Электронный ресурс]. – Режим доступа : https://www.ee.washington.edu/techsite/papers/documents/UWEETR-2004-0013.pdf
3. Lingodroids: Studies in Spatial Cognition and Language / [Schulz R., Glover A., Milford M.] // Proc. 2011 IEEE International Conference on Robotics and Automation (ICRA-2011) (May 9-13, 2011, Shanghai, China). – P. 178-183.
4. Секунов Н. Обработка звука на PC / Секунов Н. – СПб., 2001. – 1248 с.

## Literatura

1. Pavlovskij V.E. V sb. "Iskusstvennyj intellekt - problemy i perspektivy". Politehnicheskie chtenija (3 Pospelovskie cht-ja). 2008. M.: Izd. Politehnicheskogo muzeja. http://posp.raai.org/info.php?view=17 http://www.umlab.ru/index/download/evgrafov_ipm.pdf
2. Tho Nguyen and Linda G.Bushnell, Dept of EE. University of Washington. April. 2006. "Feasibility Study of DTMF Communications for Robots". https://www.ee.washington.edu/techsite/papers/documents/UWEETR-2004-0013.pdf
3. Schulz R. Proc. 2011 IEEE International Conference on Robotics and Automation (ICRA-2011). May 9-13. 2011. Shanghai. China. P. 178-183.
4. Sekunov N. "Obrabotka zvuka na PC". SPb. 2001.1248 s.

***RESUME***

***V.E. Pavlovsky, A.Ju. Kirkov***

# *Artificial Tone Language for Acoustic Communication of Robots*

The work deals with creation of system of acoustic communication of robots for realization of their conversation in the usual external (air) environment on the typical acoustic frequencies perceived by the human. Generally speaking, low-frequency language of robots won't be directly clear to the person, but as for the person it is simple to prepare the corresponding translator. Thus, offered acoustic communication can become the convenient means of additional control of behavior of the robots coexisting with the human beings in uniform society.

In the work, communication of robots is based on simple structure of one- and low-frequency signals that does methods of the spectral analysis almost ideal for the subsequent algorithms of recognition. On the basis of such signals, base language of robots is under construction. The language can include any number of the elementary symbols realized by the mono- or polyfrequency of artificial phonemes. In the work, the formal model of the language is entered, the program environment for visualization and the analysis of work of the constructed algorithm is realized.

Experiments on the basis of the created computer and hardware models for the analysis of operability of the constructed algorithms are made. The first prototype of described system was created in 2007-2008. During the subsequent time, experiments on creation of several various realization and language working off were carried out. The experiments showed the sufficient efficiency of the created acoustical system.

*Статья поступила в редакцию 16.07.2012*